How Diverse Is Your Team? Investigating Gender and Nationality Diversity in GitHub Teams

Marco Ortu[1], Giuseppe Destefanis[2], Steve Counsell[2], Stephen Swift[2], Roberto Tonelli[1], and Michele Marchesi[1]

[1] DIEE, University of Cagliari,
|marco.ortu|michele|roberto.tonelli|@diee.unica.it
[2] Brunel University, Uxbridge, UK
|giuseppe.destefanis|steve.counsell|stephen.swift|@brunel.ac.uk

Abstract. Building an effective team of developers is a complex task faced by both software companies and open source communities. The problem of forming a "dream" team involves many variables, including consideration of human factors, and it is not a dilemma solvable in a mathematical way. Empirical studies might provide interesting insights to explain which factors need to be taken into account in building a team of developers and which levers act to optimise collaboration and productivity among developers. In this paper, we present the results of an empirical study aimed at investigating the link between team diversity (i.e., gender, nationality) and productivity (issue fixing time). We consider issues solved from the GHTorrent dataset inferring gender and nationality of each team's members. We also evaluate the politeness of all comments involved in issue resolution. Results show that higher gender diversity is linked with a lower team average issue fixing time and that nationality diversity is linked with lower team politeness.

Key words: Affective Analysis, Issue Report, Empirical Software Engineering

1 Introduction

Diversity in working teams has been studied in several research fields [15, 27] and is considered as any attribute which differentiates people [31] such as demographic attributes (e.g., age, gender, nationality), functional (e.g., role, tenure, expertise), or subjective (e.g., personality). Previous research reports contrasting evidence about the role of diversity in team work: some studies report significant positive correlations between diversity and performance [11], while others found that diversity negatively impacts team outcomes [30].

Herring [14] used data from the 1996 to 1997 National Organizations Survey, to test eight hypotheses derived from the value-in-diversity thesis. The results supported seven of these hypotheses: racial diversity was associated with increased sales revenue, more customers, greater market share, and greater relative profits. Gender diversity was associated with increased sales revenue, more customers, and greater relative profits.

As far as open source development is concerned, Daniel et al. [9] studied effects of diversity on community engagement and market success in a sample of 357 SourceForge projects. Results showed that reputation and role diversity were positively correlated with market success and community engagement. Conversely, diversity of spoken languages and nationality were negatively associated with community engagement and a positive impact on market success.

Diversity in experience and language was studied by Chen et al. [7] using Wikipedia Projects. The authors examined the effects of group diversity on the amount of work accomplished and on member withdrawal behaviours in the context of WikiProjects. They found that increased diversity in experience with Wikipedia increased group productivity and decreased member withdrawal up to a threshold. Beyond that threshold, group productivity remained high, but members were more likely to withdraw.

Vasilescu et al. [28] studied gender and tenure diversity in GitHub teams and found that they were positive and significant predictors of productivity. Lin et al. [18] evaluated the applicability of different gender guessing approaches on several datasets derived from Stack Overflow. Compared to Vasilescu et al. [29], this study focused on diversity in gender and nationality and studied development teams based on issue collaboration rather than commits activity.

In this paper, we investigate the impact of gender and nationality diversity on the productivity and collaboration quality of a team. Based on evidence of previous research [19, 21, 22, 25], we used the issue fixing time and politeness as proxy metrics for the team productivity and level of collaboration, respectively. We present two logistic regression representing the average time required to solve an issue by development teams and the communication level measured by the politeness of a team. We exploit a dataset extracted from the 2014 dump of the GHTorrent dataset [12]. A set of heuristics was used to infer development teams based on GitHub's issue collaboration graph, its user's gender and nationality with the final goal of building a representative diversity dataset. We explore the following research questions:

RQ1: Are gender or nationality diversity linked to the issue fixing time of a team?
Gender diversity in GitHub teams was linked with lower Issue Fixing Time (IFT).
Our model showed that gender diversity was the most dominant metric explaining data variance and was positively associated with productivity (we observed shorter issue fixing time in teams with higher gender diversity).

RQ2: Are gender or nationality diversity linked to the overall politeness of a team?
Country diversity was linked with lower politeness.
Our model showed that country diversity was the most dominant metric ex-

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2285v1 | CC BY 4.0 Open Access | rec: 13 Jul 2016, publ: 13 Jul 2016

plaining data variance and had a negative effect on team politeness (it had a tendency to lower politeness).

The rest of the paper is organised as follows: in Section 2, we describe how we measured politeness and how issue collaboration graphs were built. Section 3 introduces the experimental design; in Section 4 we present the case study setup, while in Section 5 we present and discuss our findings, followed by a discussion of threats to validity in Section 6. Finally, we draw conclusions in Section 7.

2 Background

In this section, we provide background and motivation about the framework we adopted in our analysis. Politeness along with other affective metrics, i.e., sentiment and emotions, has been used in several studies in software engineering. [10, 13, 21, 22, 23, 24, 26]. In the following, we briefly introduce politeness and the methodology used to infer GitHub development teams.

2.1 Politeness

Politeness is "the ability to make all the parties relaxed and comfortable with one another[1]." Politeness is crucial in collaboration as people tend to perceive linguistic markers of politeness as a form of respect and their use is related to the power dynamics of social interaction. Danescu et al. [8] performed an empirical study on Wikipedia and Stack Exchange to investigate the relation between politeness and social dynamics. They showed how polite editors in Wikipedia were more likely to achieve higher status in the community, thus suggesting a positive association between politeness and positive social outcomes. Conversely, they showed how politeness negatively correlates with social status: once elected, the Wikipedia editors became less polite; similarly, high reputation users of Stack Exchange tended to exhibit a less polite linguistic behaviour. Burke and Kraut [6] investigated the impact of politeness on community engagement, measured in terms of reply rates. They showed how politeness differently affected the reply rates based on the topic being discussed. In particular, rudeness appeared to be more effective in eliciting responses in political discussion while a polite attitude attracted more contribution in technical groups where people were typically seeking help. This was consistent with the results of the study by Althoff et al. [1] on altruistic requests in online communities, which showed that paying gratitude forward positively correlated to success, i.e. to an increased probability of receiving help.

2.2 Issue Collaboration Graph

GitHub's issue tracking is notable because it is focused on collaboration, it is a good way of keeping track of tasks, enhancements and bugs for projects and it is

[1] `http://en.wikipedia.org/wiki/Politeness`

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2285v1 | CC BY 4.0 Open Access | rec: 13 Jul 2016, publ: 13 Jul 2016

primary designed for collaboration. An issue generally consists of several parts such as:

– A title and description indicating what the issue is about.
– Colour-coded labels to help categorise and filter issues (like labels in email).
– A milestone which acts like a container for issues. This is useful for associating issues with specific features or project phases (e.g. Weekly Sprint 9/5-9/16 or Shipping 1.0).
– A reporter, who originally opened the issue.
– An assignee, responsible for working on the issue at any given time.
– Comments, which allow members of the repository to provide feedback.

Developers can post their comments on an issue report to discuss and manage the issue resolution. To build a network graph based on issue comments, we considered each developer as a node; when user A commented on an issue reported by user B, we considered it as an edge from developer A to developer B. Following this approach, we obtained a directed network graph which we called *an issue collaboration graph*. Fig. 1 shows an example of issue collaboration graph in which different colours are related to different team of developers.

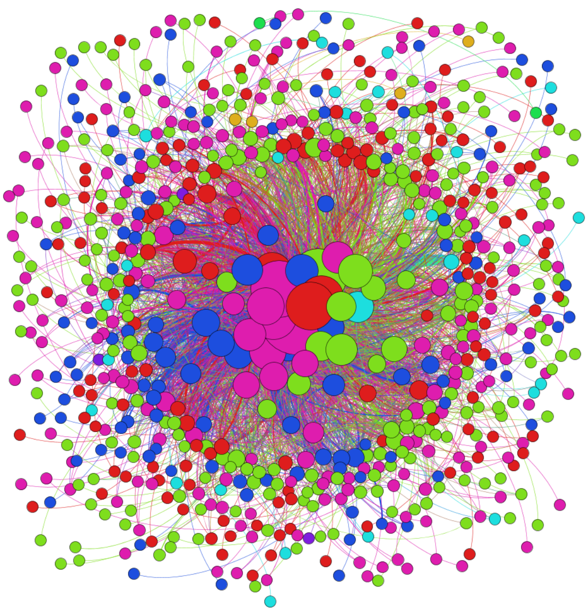


Fig. 1: Example of Issue Collaboration Graph as visualised by Gephi (Hadoop Common project).

The dimension of a node represents the amount of activity of that specific developers, e.g., bigger nodes represent more active developers within a given team. We considered users who performed at least one commit for the project, hence, generic users (allowed to open issues on GitHub) were not considered in the network. In this paper, user and developer are considered synonyms.

We then used Gephi (an interactive visualization and exploration tool [2]) to analyze the obtained network. In particular, since information about team members are not provided on GitHub (we know that all the developers in the network are involved in the same project, but we do not know how the workload is distributed among them), we ran a modularity algorithm, based on the algorithms developed by Blonde [5] and Lambiotte [17], to obtain the communities (teams in this specific case) present in the network. Blonde et al. [5] proposed a method to extract the community structure of large networks. It is a heuristic method that is based on modularity optimisation.

3 Experimental Design

We built a diversity dataset to infer gender and country of the GitHub users, extracted the issue collaboration graph (Section2.2) and performed modularity analysis to obtain the teams shown in Fig. 2.

We used the last 2014 dump of the GHTorrent dataset [12]. The dataset is a mirror of GitHub's data packed as a MySQL database. In our study, we were interested in modelling a developer network based on collaboration on issue resolution. For this purpose, only closed issues with at least 2 comments posted by different developers (including the issue reporter) were used. Finally, we considered 33673 issues, with 71423 comments posted by 13872 developers belonging to 1176 different teams as shown in Table 1.

Statistics	Value
Projects analyzed	8040
# of Developers	13872
# of Issues	33673
# of Comments	71423
# of Teams	1176
Average # of fixed issues per team	278

Table 1: Dataset Statistics.

Two logistic regression model were modelled as follows.

Response Variables

Team Productivity

We measured team productivity considering the average IFT of a team and, since we were using a logistic regression, considered the team average issue fixing time to be 0 or 1 meaning that the average IFT was higher or lower than the IFT median.

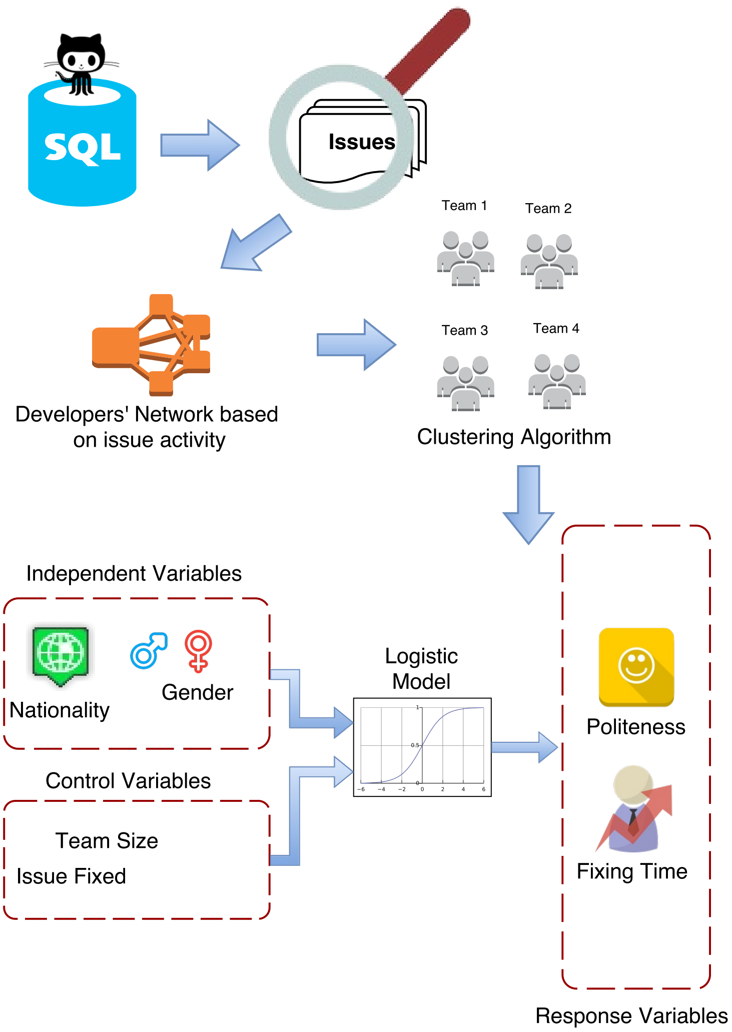


Fig. 2: Experimental design schema.

Team Politeness
Team politeness was measured as the ratio between polite and impolite comments posted on issues resolved by a team. To obtain a discrete variable, we used a logistic regression model and considered a politeness of 1 if the ratio was greater than 1 (more polite comment) and a politeness of 0 if the polite/impolite ratio was less than 1 (namely there were more impolite comments).

Indipendent Variables

Gender diversity. We measured gender diversity within a team using Blau's Diversity Index [4], $1 - \sum_1^N p_i^2$, p is a categorical variable male/female. We only considered users we could infer gender from the evaluation of the gender diversity index.

Nationality Diversity. Country diversity within a team was measured using Blau's Diversity Index [4], $1 - \sum_1^N p_i^2$, p in this case is a categorical variable indicating the country (i.e. Italy, UK, etc); only users we could infer the country in the evaluation of the nationality diversity index were considered.

Control Variables

Team Size. Other studies have suggested that the total number of developers involved in an issue resolution is linked with longer IFTs and less polite comments [22, 24]; we thus considered the total number of developers in a team as a control variable for our experiment.

Team tenure. Team tenure is intended as the number of issues resolved by a given team. It was considered as a control variable since developer experience is linked with both IFT and politeness [22, 24]. Average IFT plays a dual role in our models. It is considered as a response variable in the productivity model and as control variable in the politeness model (a longer IFT is linked with a lower politeness [22]).

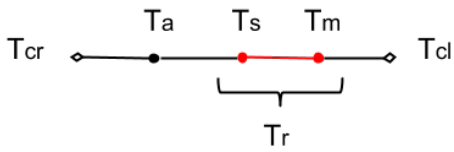
4 Case Study Setup

4.1 Measuring the Control Variables

Inferring Gender Following the approach and tool provided by Vasilescu et al. [28] we inferred the GitHub user's gender. Combining heuristics with female/-male frequency name lists collected and the country the users belong to, this tool is able to infer gender. Country data is crucial for inferring a user's gender from their name (e.g. Andrea is a common male name in Italy, but a common female name in many other countries). The reported precision of gender is 93% [28]. In order to infer user's gender, it is necessary to know their name and country.

Inferring User Name And Country The simple task of retrieving the first name of a GitHub user is challenging due to the noise and lack of information in the dataset [16]. Approximately 20% of all users were labeled as "unknown" in the GHTorrent dump analysed. Approximately 9% of all users have at least two accounts. Thus, we grouped together all user's accounts with a set of heuristics proposed by Bird et al. [3], i.e. accounts with same email were merged. We inferred user nationality by applying a set of heuristics [2] based on the *location* available data; we were able to determine name and country of about 31% of all users, a percentage close to that obtained in similar studies [28, 29].

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2285v1 | CC BY 4.0 Open Access | rec: 13 Jul 2016, publ: 13 Jul 2016



$$Tr = Tm - Ts \tag{1}$$

Fig. 3: Example of timeline for GitHub issue.

4.2 Measuring the Dependent Variables

Inferring Issues Fixing Time Figure 3 shows the typical issue timeline in GitHub:

- T_{cr} represents the time when an issue is created.
- T_{cl} represents the time when an issue is closed.
- T_{a} represents the time when an issue is assigned to a developer.
- T_{s} is the time when a developer subscribes to an issue that has been assigned to them.
- T_{m} represents the time when an issue is merged with the repository, namely the local commit is merged with the remote repository.

To infer the issue fixing time (abbreviated as IFT), we used the approach proposed by Murgia et al. [20]. We computed the time interval between the last time an issue had been merged and the last time it had been subscribed to by an assignee (issues and pull requests are dual on Github, for each opened pull request, an issue is opened automatically [12]). In the case we could not obtain such dates, we used a conservative approach (i.e., if the subscribed date was missing we used assigned data and so on).

Measuring Politeness To compute the polarity of the contribution in our dataset, we adopted the library developed by Danescu et al. [8]. Given an input text, the tool calculates its overall politeness in for of a discrete label, i.e. *polite* or *impolite*.

The tool was trained and validated through machine learning on a gold standard of over 10,000 manually labeled requests from Wikipedia[3] and Stack Overflow[4]. The gold standard was built so as to include comments written by authors from all over the world while the annotators were selected among U.S. residents, based on a linguistic background questionnaire to reduce. The classifiers has been evaluated both in an in-domain setting, with a standard leave-one-out cross validation procedure, and in a cross-domain setting, where they trained

[2] https://github.com/tue-mdse/countryNameManager
[3] `https://en.wikipedia.org/wiki/Main_Page`
[4] `http://stackoverflow.com`

on one domain and tested on the other Danescu et al. [8], achieving accuracy of 78.19% and 75.43%, respectively. Based on this evidence, we considered the tool by Danescu et al. [8] robust enough to be adopted in our domain, i.e., Jira[5] issues, where developers post and discuss about technical aspects of issues.

5 Results

To conduct our analysis, we applied a logistic regression for estimating the extent to which gender and nationality diversity influenced productivity and the level of collaboration in a team. We used logistic regression for its ease of interpretability, since it allowed us to reason about the significance of one factor given all the others. The results of the logistic regression are reported in Table 2 and 3. The tables list the results for each of the predictors in our framework, that is for the independent variables (namely, nationality and gender) and the control variables namely, team size and issue and the team tenure) defined in our framework, grouped by actionable factor. For each predictor, logistic regression outputs three values, namely coefficient estimate, odds ratio and statistical significance. The sign of the coefficient estimate indicates the positive/negative impact of the predictor on the success of a question. The odds ratio weighs the magnitude of this impact: the closer the value is to 1, the smaller the impact of the parameter on the chance of success. In particular, an odds ratio value lower than 1 corresponds to a negative impact of the predictor, and vice versa. Finally, the statistical significance determines whether a predictor has a significant explanatory value.

5.1 RQ1: Are gender or nationality diversity linked to the issue fixing time of a team?

Motivation. When building a team, one of the first goals is a high level of effectiveness. While there are many ways for measuring effectiveness of a working team, i.e, number of activities solved in a time interval or the total time to finish all activities, we considered a measure of productivity as the average time required to resolve an issue. Knowing which factor could effect this measure is crucial for a successful team.
Approach. Using the methodology described in Section 4, we assembled a longitudinal dataset of GitHub teams based on the issue collaboration graph. We modelled team productivity as the average IFT and considered gender and country diversity as team characteristics controlled by the team size and number of issues solved.
Findings. Gender diversity in GitHub teams was linked with lower IFT.
In our model, we modelled average IFT as a binary variable where 1 means a value higher than the median (considering all issues) and 0 means lower. The model showed that gender diversity was the most dominant metric explaining

[5] `https://www.atlassian.com/software/jira`

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2285v1 | CC BY 4.0 Open Access | rec: 13 Jul 2016, publ: 13 Jul 2016

data variance and had a negative effect on productivity (tendency to lower issue fixing time, Table 2). Country diversity, along with the other control metrics, were not significant. We accept that this is a first study on diversity in GitHub teams; more sophisticated models are needed to confirm our findings.
Our finding matches common sense. It is expected that diversity brings novelty, unique experience and different perspectives which can explain a raise in term of productivity.

Coefficients	Estimate	z value	Odd's Ratio	Pr(> $\|z\|$)
(Intercept)	0.61	0.604	1.83	0.54
Team tenure	-0.0001	-0.882	0.9	0.37
team size	0.0007	1.195	1	0.23
gender div	-2.72	-2.709	0.06	**
country div	0.005	0.005	1	0.99

Table 2: Issue Fixing Time Model. Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

5.2 RQ2: Are gender or nationality diversity linked to the overall politeness of a team?

Motivation. The level of collaboration within a team is as important as productivity, since poor communication and collaboration lead to unproductive and unsuccessful teams. Studying a team's collaboration is crucial for highlighting which factors have an influence. Measuring the overall politeness expressed by a development team while discussing an issue resolution, we have been able to obtain useful insights into which factors might affect the overall team collaboration.
Approach. We assembled a longitudinal dataset of GitHub teams based on the issue collaboration graph and modelled team politeness ratio as the model outcome. We considered gender and country diversity as team characteristics controlled by the team size, number of issues solved and average IFT.
Findings. Country diversity is linked with lower politeness.
We modelled team politeness as a binary variable. A value of 1 means that the ratio polite/impolite comments is greater than 1 (more polite comments) and 0 means less than 1 (more impolite comments). The model shows that country diversity is the most dominant metric explaining data variance and has a negative effect on team politeness (it has a tendency to lower politeness). Gender diversity, along with the other control metrics, is not significant in our model (Table 3).
Lower politeness within a team of "diverse" developers in term of nationality could be related to habits and customs co-existing in the same environment. We analysed teams of developers of open source projects, hence, people working remotely and communicating through emails, chat, short messages exchanged on

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2285v1 | CC BY 4.0 Open Access | rec: 13 Jul 2016, publ: 13 Jul 2016

issue tracking systems. The open source paradigm brings constraints due to lack of shared physical spaces and personal bonds. Different language backgrounds may be a prolific source of misunderstanding and misinterpretation of written requests or comments, leading to a general lower level of politeness.

Coefficients	Estimate	z value	Odd's Ratio	Pr(> $\|z\|$)
(Intercept)	4.46	0.914	0.15	0.36
Team tenure	-1.35e-3	-1.451	0.9	0.14
team size	2.3e-3	1.261	1	0.20
gender div	-5.57	-0.579	5e-3	0.56
country div	-12.48	-2.074	1	*
fixing time	5.15e-7	1.774	1	0.12

Table 3: Politeness Model.Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

6 Threats To Validity

Threats to internal validity concern confounding factors that can influence the obtained results. We assume a causal relationship between a developer's emotional state and what they write in issue report comments, based on empirical evidence (in another domain). We built an explanatory model to understand the characteristics of development teams considering productivity and collaboration.

Threats to construct validity focus on how accurately the observations describe the phenomena of interest. We used state-of-the-art tools [8] provided by Danescu et al. to measure politeness, in addition to heuristics for evaluating issue fixing time.

Threats to external validity correspond to the generalisability of our experimental results. We consider issues as a representative sample of the universe of open source software projects, with different development teams and satisfying different customers' needs. Replications of this work on other open source systems and on commercial projects are needed to confirm our findings.

Other threats concern the validity of the models used for our analysis. In particular, the models presented are probably not complex enough to properly represent the measured phenomena (as suggested by the low significance of *control variables* in both experiments).

7 Conclusions

In this paper, we presented two logistic regression models representing the average time required to solve an issue by development teams and the overall communication level measured by the overall politeness of a team. Results showed that gender diversity is linked with lower average issue fixing time and nationality diversity is linked with lower team politeness. We used a set of heuristics to infer development teams based on GitHub's issues collaboration graph, GitHub's user gender and nationality with the final goal of building a representative diversity dataset. Our results also indicated that both gender and nationality diversity played a significant role when considering the productivity and collaboration within a team. We believe that these results are a good starting point for stimulating more research activities toward this direction and could help managers and team directors in taking better decisions during the crucial starting phase when building a team of developers.

8 Acknowledgement

The research presented in this paper was partly funded by the Engineering and Physical Sciences Research Council (EPSRC) of the UK under grant ref: EP/M024083/1.

References

1. T. Althoff, C. Danescu-Niculescu-Mizil, and D. Jurafsky. How to ask for a favor: A case study on the success of altruistic requests. In *Proceedings of ICWSM*, ICWSM '14, 2014.
2. M. Bastian, S. Heymann, M. Jacomy, et al. Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8:361–362, 2009.
3. C. Bird, A. Gourley, P. Devanbu, M. Gertz, and A. Swaminathan. Mining email social networks. In *Proceedings of the 2006 international workshop on Mining software repositories*, pages 137–143. ACM, 2006.
4. P. M. Blau. *Inequality and heterogeneity: A primitive theory of social structure*, volume 7. Free Press New York, 1977.
5. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
6. M. Burke and R. Kraut. Mind your ps and qs: The impact of politeness and rudeness in online communities. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, CSCW '08, pages 281–284, New York, NY, USA, 2008. ACM.
7. J. Chen, Y. Ren, and J. Riedl. The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 821–830. ACM, 2010.
8. C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, and C. Potts. A computational approach to politeness with application to social factors. In *Proceedings of ACL*, 2013.
9. S. Daniel, R. Agarwal, and K. J. Stewart. The effects of diversity in global, distributed collectives: A study of open source project success. *Information Systems Research*, 24(2):312–333, 2013.

10. G. Destefanis, M. Ortu, S. Counsell, S. Swift, M. Marchesi, and R. Tonelli. Software development: do good manners matter? *PeerJ Computer Science*, 2016.
11. C. P. Earley and E. Mosakowski. Creating hybrid team cultures: An empirical test of transnational team functioning. *Academy of Management Journal*, 43(1):26–49, 2000.
12. G. Gousios. The ghtorrent dataset and tool suite. In *Proceedings of the 10th Working Conference on Mining Software Repositories*, MSR '13, pages 233–236, Piscataway, NJ, USA, 2013. IEEE Press.
13. E. Guzman, D. Azócar, and Y. Li. Sentiment analysis of commit comments in github: an empirical study. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 352–355. ACM, 2014.
14. C. Herring. Does diversity pay?: Race, gender, and the business case for diversity. *American Sociological Review*, 74(2):208–224, 2009.
15. S. K. Horwitz and I. B. Horwitz. The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of management*, 33(6):987–1015, 2007.
16. E. Kalliamvakou, G. Gousios, K. Blincoe, L. Singer, D. M. German, and D. Damian. The promises and perils of mining github. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 92–101. ACM, 2014.
17. R. Lambiotte, J.-C. Delvenne, and M. Barahona. Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*, 2008.
18. B. Lin and A. Serebrenik. Recognizing gender of stack overflow users. In *Proceedings of the 13th International Workshop on Mining Software Repositories*, pages 425–429. ACM, 2016.
19. M. Mäntylä, B. Adams, G. Destefanis, D. Graziotin, and M. Ortu. Mining valence, arousal, and dominance: possibilities for detecting burnout and productivity? In *Proceedings of the 13th International Workshop on Mining Software Repositories*, pages 247–258. ACM, 2016.
20. A. Murgia, G. Concas, R. Tonelli, M. Ortu, S. Demeyer, and M. Marchesi. On the influence of maintenance activity types on the issue resolution time. In *Proceedings of the 10th International Conference on Predictive Models in Software Engineering*, pages 12–21. ACM, 2014.
21. A. Murgia, P. Tourani, B. Adams, and M. Ortu. Do developers feel emotions? An exploratory analysis of emotions in software artifacts. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 262–271. ACM, 2014.
22. M. Ortu, B. Adams, G. Destefanis, P. Tourani, M. Marchesi, and R. Tonelli. Are bullies more productive? Empirical study of affectiveness vs. issue fixing time. In *Proceedings of the 12th Working Conference on Mining Software Repositories*, MSR 2015, 2015.
23. M. Ortu, G. Destefanis, S. Counsell, S. Swift, R. Tonelli, and M. Marchesi. Arsonists or firefighters? Affectiveness in agile software development. In *International Conference on Agile Software Development*, pages 144–155. Springer, 2016.
24. M. Ortu, G. Destefanis, M. Kassab, S. Counsell, M. Marchesi, and R. Tonelli. Would you mind fixing this issue? An empirical analysis of politeness and attractiveness in software developed using agile boards. In *Agile Processes, in Software Engineering, and Extreme Programming*, pages 129–140. Springer, 2015.
25. M. Ortu, G. Destefanis, M. Kassab, and M. Marchesi. Measuring and understanding the effectiveness of JIRA developers communities. In *Proceedings of the Sixth International Workshop on Emerging Trends in Software Metrics*, pages 3–10. IEEE Press, 2015.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2285v1 | CC BY 4.0 Open Access | rec: 13 Jul 2016, publ: 13 Jul 2016

26. D. Pletea, B. Vasilescu, and A. Serebrenik. Security and emotion: sentiment analysis of security discussions on GitHub. In *Proceedings of the 11th Working Conference on Mining Software Repositories*, pages 348–351. ACM, 2014.
27. G. K. Stahl, M. L. Maznevski, A. Voigt, and K. Jonsen. Unraveling the effects of cultural diversity in teams: A meta-analysis of research on multicultural work groups. *Journal of international business studies*, 41(4):690–709, 2010.
28. B. Vasilescu, A. Capiluppi, and A. Serebrenik. Gender, representation and online participation: A quantitative study. *Interacting with Computers*, 2013.
29. B. Vasilescu, D. Posnett, B. Ray, M. G. van den Brand, A. Serebrenik, P. Devanbu, and V. Filkov. Gender and tenure diversity in GitHub teams. In *CHI. ACM*, 2015.
30. W. E. Watson, K. Kumar, and L. K. Michaelsen. Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of management journal*, 36(3):590–602, 1993.
31. K. Y. Williams and C. A. O'Reilly. Demography and diversity in organizations: A review of 40 years of research. *Research in organizational behavior*, 20:77–140, 1998.